

AD-A170 306

EXPLOITING AUXILIARY INFORMATION ABOUT EXAMINEES IN THE  
ESTIMATION OF ITEM PARAMETERS(U) EDUCATIONAL TESTING  
SERVICE PRINCETON NJ R J MISLEVY MAY 86

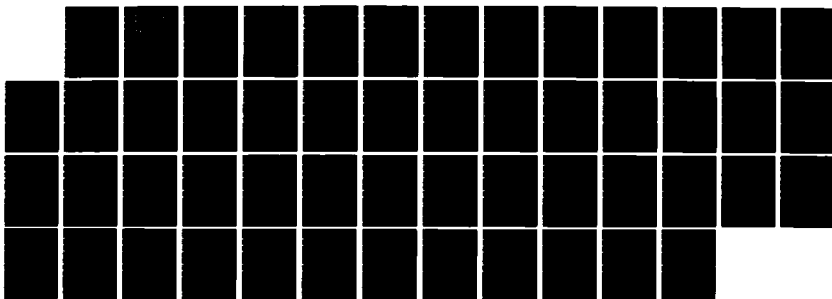
1/1

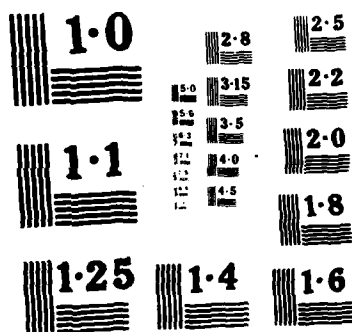
UNCLASSIFIED

ETS-RR-86-18-ONR N00014-85-K-0683

F/G 12/1

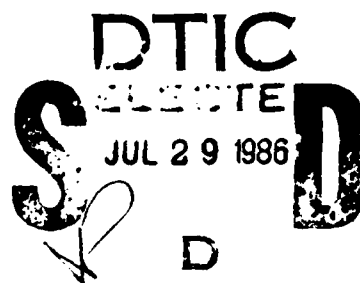
NL





12

AD-A170 306

**RESEARCH****REPORT**

# EXPLOITING AUXILIARY INFORMATION ABOUT EXAMINEES IN THE ESTIMATION OF ITEM PARAMETERS

Robert J. Mislevy

This research was sponsored in part by the  
Personnel and Training Research Programs  
Psychological Sciences Division Office of Naval  
Research, under Contract No. N00014-85-K-0683

Contract Authority Identification Number NR  
No. 150-539

Robert J. Mislevy, Principal Investigator



Educational Testing Service  
Princeton, New Jersey

May 1986

Reproduction in whole or in part is permitted for  
any purpose of the United States Government.

Approved for public release; distribution  
unlimited.

DTIC FILE COPY

86 7 28 170

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

AD-A170306

## REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified		1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.	
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE			
4. PERFORMING ORGANIZATION REPORT NUMBER(S) RR-86-18-ONR		5. MONITORING ORGANIZATION REPORT NUMBER(S)	
6a. NAME OF PERFORMING ORGANIZATION Educational Testing Service	6b. OFFICE SYMBOL (if applicable)	7a. NAME OF MONITORING ORGANIZATION Personnel and Training Research Programs Office of Naval Research	
6c. ADDRESS (City, State, and ZIP Code) Princeton, NJ 08541		7b. ADDRESS (City, State, and ZIP Code) Code 442PT 800 North Quincy Street Arlington, VA 22217-5000	
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Office of Naval Research	8b. OFFICE SYMBOL (if applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-85-K-0683	
8c. ADDRESS (City, State, and ZIP Code) 800 North Quincy Street Arlington, VA 22217-5000		10. SOURCE OF FUNDING NUMBERS	
		PROGRAM ELEMENT NO 61153N	PROJECT NO RR04204
		TASK NO RR04204-01	WORK UNIT ACCESSION NO 4421539
11. TITLE (Include Security Classification) Exploiting Collateral Information in the Estimation of Item Parameters (Unclassified)			
12. PERSONAL AUTHOR(S) Mislevy, Robert J.			
13a. TYPE OF REPORT Technical Report	13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Year, Month, Day) May 1986	15. PAGE COUNT 42
16. SUPPLEMENTARY NOTATION			
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB-GROUP	
19. ABSTRACT (Continue on reverse if necessary and identify by block number)			
<p>The precision of item parameter estimates can be increased by taking advantage of dependencies between the latent proficiency variable and auxiliary examinee variables such as age, courses taken, and years of schooling. Gains roughly equivalent to two to six additional item responses can be expected in typical educational and psychological applications. Empirical Bayes computational procedures are presented, and illustrated with data from the <u>Profile of American Youth</u> survey.</p>			
20. DISTRIBUTION / AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS		21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Robert J. Mislevy		22b. TELEPHONE (Include Area Code) (609) 734-1271	22c. OFFICE SYMBOL

Exploiting Auxiliary Information

1

EXPLOITING AUXILIARY INFORMATION ABOUT EXAMINEES IN THE  
ESTIMATION OF ITEM PARAMETERS

Robert J. Mislevy

This research was sponsored in part by the  
Personnel and Training Research Programs  
Psychological Sciences Division  
Office of Naval Research, under  
Contract No. N00014-85-K-0683

Contract Authority Identification Number  
NR No. 150-539

Robert J. Mislevy, Principal Investigator

Educational Testing Service  
Princeton, New Jersey

May 1986

Reproduction in whole or in part is permitted  
for any purpose of the United States Government.

Approved for public release; distribution  
unlimited.

DISTRIBUTION	
Availability Codes	
Dist	Avail and/or Special
A-1	

Abstract

The precision of item parameter estimates can be increased by taking advantage of dependencies between the latent proficiency variable and auxiliary examinee variables such as age, courses taken, and years of schooling. Gains roughly equivalent to two to six additional item responses can be expected in typical educational and psychological applications. Empirical Bayes computational procedures are presented, and illustrated with data from the Profile of American Youth survey.

Key words: EM-algorithm, empirical Bayes, marginal maximum likelihood

Exploiting Auxiliary Information about Examinees in the  
Estimation of Item Parameters

A pervasive problem in item response theory (IRT) is the difficulty of simultaneously estimating large numbers of parameters from limited data. Even large samples of examinees may not eliminate the problem when each examinee responds to only a few items, as in educational assessment and adaptive testing. Certain improvements are obtained by using hierarchical models along the lines of Lindley and Smith (1972); treating examinee parameters as a sample from a common population enhances the stability and precision of item parameter as well as examinee parameter estimates. This approach has been applied to IRT by a number of researchers recently, including Bock and Aitkin (1981), Leonard and Novick (1985), Rigdon and Tsutakawa (1982), and Swaminathan and Gifford (1982).

For the most part, the aforementioned writers consider all examinees to be members of a single, undifferentiated, population. This framework instantiates such beliefs as, "if the parameters of most examinees seem to lie between  $-3$  and  $+3$ , then the parameter of an examinee who answered both of two hard math items correctly is probably somewhere between  $+1.5$  to  $+3.5$ --even though his/her maximum likelihood estimate is  $+\infty$ ." Additional stability and precision may yet be achieved if auxiliary information is

available about examinees, such as educational background or status on demographic variables. A statement like "the parameter of an examinee who answered both of two hard math items correctly and studied calculus in college is probably between +2.7 and +3.7," might result.

This paper addresses the utilization of auxiliary information about examinees in estimating item parameters. The following section reviews item parameter estimation when examinee parameters are known, then when examinee parameters are unknown and nothing is assumed about them. Attention then turns to the additional assumptions of first, an undifferentiated population, and second, a population differentiated with respect to auxiliary variables. Following this are sections that discuss anticipated gains in precision, outline computational procedures, and illustrate the approach with responses to four items from the Arithmetic Knowledge subtest of the Armed Services Vocational Aptitude Battery.

#### The Role of Auxiliary Information

The relevance of auxiliary examinee variables to item parameter estimation is not immediately obvious, since they play no role in the basic model for item responses. Letting  $\underline{x}_i = (x_{i1}, \dots, x_{in})$  represent the responses of examinee  $i$  to  $n$  test items and  $y_i$  represent values of auxiliary variables such as educational and demographic status, the standard IRT assumption of local



independence states that

$$p(\underline{x}_i | \theta_i, y_i, \underline{\beta}) = \prod_j^n p(x_{ij} | \theta_i, \beta_j) \quad , \quad (1)$$

where  $\theta_i$  is the examinee parameter,  $\underline{\beta} = (\beta_1, \dots, \beta_n)$  are possibly vector-valued item parameters, and the form of  $p(x_{ij} | \theta_i, \beta_j)$  is specified a priori through the item response model. It follows that  $y_i$  would indeed be irrelevant to item parameter estimation if  $\theta_i$  were known. The likelihood to be maximized with respect to  $\underline{\beta}$ , given the data matrix  $X = (\underline{x}_1, \dots, \underline{x}_N)$  of responses from  $N$  examinees with proficiencies  $\underline{\theta} = (\theta_1, \dots, \theta_N)$  and auxiliary variables  $\underline{Y} = (y_1, \dots, y_N)$ , would be simply

$$L = \prod_i^N p(\underline{x}_i | \theta_i, \underline{\beta}) \quad . \quad (2)$$

The maximum likelihood estimate (MLE)  $\hat{\underline{\beta}}$  would satisfy the likelihood equations

$$\underline{0} = \sum_i \partial \ell_i(\theta_i) / \partial \underline{\beta} \quad , \quad (3)$$

where  $\ell_i(\theta) = \log p(\underline{x}_i | \theta, \underline{\beta})$ , and the covariance matrix of

estimation error variances for  $\hat{\beta}$  could be approximated by the inverse of the observed information matrix  $I$ :

$$I_{\theta} = \sum_i \left( \frac{\partial \ell_i(\theta_i)}{\partial \beta} \right) \left( \frac{\partial \ell_i(\theta_i)}{\partial \beta'} \right) \bigg|_{\beta=\hat{\beta}} \quad (4)$$

But Equation 1 gives response probabilities conditioned on  $\theta$ , and  $\theta$  is not known in practice. The problem that must actually be solved is to maximize the marginal likelihood

$$L_M = \prod_i \int p(x_i | \theta, \beta) dF_i(\theta) \quad (5)$$

where  $F_i(\theta)$  is the distribution of the unknown proficiency of examinee  $i$ . This is an "incomplete data" problem, in the terminology of Dempster, Laird, and Rubin (1977), corresponding to the "complete data" problem of maximizing Equation 2 when  $\theta$  is known. Assuming the required integrals exist, the likelihood equations become

$$0 = \sum_i p_i^{-1}(x_i) \int [\partial \ell_i(\theta) / \partial \beta] dF_i(\theta) \quad ,$$

where

$$p_i(\underline{x}_i) = \int \ell_i(\theta) dF_i(\theta) \quad .$$

Louis (1982) shows that if Zack's (1971, Chapter 5) regularity conditions are met and if  $F_i$  is known for all  $i$ , the diagonal elements of the incomplete-data observed information matrix, namely

$$I_x = \sum_i p_i^{-1}(\underline{x}_i) \int \left\{ \left( \frac{\partial \ell_i(\theta)}{\partial \underline{\beta}} \right) \left( \frac{\partial \ell_i(\theta)}{\partial \underline{\beta}'} \right) \right\} \bigg|_{\underline{\beta}=\hat{\underline{\beta}}} dF_i(\theta) \quad , \quad (6)$$

cannot exceed the diagonal elements  $I_{\theta}$ . In other words, the precision with elements of  $\hat{\underline{\beta}}$  would be estimated if  $\underline{\theta}$  were known provides an upper limit to the precision to be expected when  $\underline{\theta}$  is not known but must be inferred.

A similar phenomenon arises in the context of sample survey analysis when a clustered sampling design is employed to estimate a mean. If  $n$  units are sampled from each of  $N$  randomly-selected clusters, then the squared standard error of the mean, ignoring finite population corrections, is given as

$$SEM^2 = \frac{\sigma^2}{nN} [1 + (n - 1)\rho] \quad ,$$

where  $\sigma^2$  is the population variance and  $\rho$  is the intraclass correlation coefficient indicating within-cluster homogeneity. If the number of clusters ( $N$ ) is held constant, increasing the sample size ( $n$ ) within clusters cannot decrease  $SEM^2$  below  $\rho\sigma^2/N$ , the value of  $SEM^2$  obtained when the means of the sampled clusters are known without error.

The estimation of  $\beta$  in the context of IRT must also deal with uncertainty from two sources. First is the usual limitation of having data from only a finite sample of examinees. All other conditions remaining unchanged, increasing  $N$  leads to greater precision for  $\hat{\beta}$ . Second is the limitation that  $\theta$  remains unknown even for sampled examinees. For a fixed sample of examinees, reducing uncertainty about  $\theta$  leads to greater precision for  $\hat{\beta}$ . This can be achieved through (i) item responses, (ii) assumptions about the  $F_1$ 's and (iii) auxiliary variables related to  $\theta$ .

de Leeuw and Verhelst (1984) point out that finding maxima in terms of  $\beta$  and of each individual  $\theta_i$  in the manner suggested by Birnbaum (1968) is equivalent to maximizing Equation 5 when each  $F_1$  concentrates its mass at the single (unknown) point  $\theta_i$ . This joint maximum likelihood (JML) solution utilizes only information in responses  $x_i$  from examinee  $i$  to reduce uncertainty about  $\theta_i$ .

Alternatively, one may consider the  $\theta$ 's to be identically distributed, so that  $F_i = F$  for all  $i$ . An auxiliary variable  $y$  is thereby implied for all examinees, an indicator signifying that each is a member of the population whose distribution is specified by  $F$ . Appearing in the literature are treatments that assume a completely specified form for  $F$  (e.g., Bock & Lieberman, 1970), others that assume parametric forms with unknown parameters  $\alpha$  to be estimated along with  $\beta$  (e.g., Zwarts & Veldhuesen, 1985), and still others that provide nonparametric approximations (e.g., Tjur, 1982). Under the first of these three approaches, the assumed population distribution combines with  $x_i$  to produce  $p(\theta_i | X)$ , which in this case equals  $p(\theta_i | x_i)$ . Under the latter two approaches, responses from examinees other than examinee  $i$  also play a role in estimating  $F$  so that  $p(\theta_i | x_i) \neq p(\theta_i | X)$ .

A third alternative, falling between unique, unconstrained  $F_i$ 's and identical  $F_i$ 's, is to posit distributions that depend on auxiliary variables: that is,  $F_i(\theta) = F_{y_i}(\theta)$ . Examinees with identical  $y$  values are considered a random sample from a population indexed by that particular value of  $y$ , and these conditional distributions are allowed to vary with  $y$ . A following section gives details for two special cases, namely a linear model and a (quasi-) nonparametric mixture approximation.

## How Much Can Be Gained?

Several factors contribute to the magnitude of the precision gains that can be achieved through population assumptions and auxiliary variables. One factor is the sensitivity of different model parameters to missing information. Mislevy's (1984) analysis of Bock and Lieberman's (1970) LSAT data showed that estimates of the population variance were more substantially improved by increases in test length than were estimates of the population mean. This might lead one to expect increased information about  $\theta$  to have more effect on item slopes than on item thresholds in the context of item parameter estimation.

A second factor is the nature of the joint distribution of auxiliary variables with  $\theta$ . An auxiliary variable adept at identifying low proficiency examinees, for example, adds information for those examinees most useful for estimating lower asymptote item parameters.

A third factor is the dependence of the estimated information upon estimated parameter values. Although a slope parameter may be consistently estimated under both the undifferentiated and undifferentiated population models, a higher estimate under the latter may appear less precise. This is because estimated standard errors for slopes are directly proportional to the values of the slope estimates, even though true standard errors depend on true

slope values and not their estimates. A slope estimated with the aid of auxiliary variables and obtaining a higher estimate can thus have a lower true standard error but a higher estimated standard error.

Since the same factors determine information gain from both increased test length and auxiliary variables, however, it is reasonable to consider the contribution of auxiliary variables in units of additional item responses. In the special case of dichotomous items, the amount of information conveyed by item responses alone is

$$i(\theta) = \sum_j \frac{P_j'(\theta)^2}{P_j(\theta)[1 - P_j(\theta)]} ,$$

where  $P_j(\theta) = p(x_j = 1|\theta)$  and  $P_j'(\theta) = dP_j(\theta)/d\theta$ . For examinees with finite maximum likelihood estimates, Bayes theorem applied with a diffuse prior leads to the approximation  $p(\theta|x_1) \approx N(\hat{\theta}, \sigma_x^2)$  with  $\sigma_x^2 = i^{-1}$ . This follows by first rescaling the likelihood so that it integrates to one, then using its mode and curvature at the mode in a normal approximation.

Consider as an example the two-parameter logistic model, under which  $P_j(\theta) \equiv p(x = 1|\theta, a_j, b_j) = 1/\{1 + \exp[-1.7a_j(\theta - b_j)]\}$ .

The contribution of item  $j$  to information about  $\theta$  is  $2.89a_j^2 P_j(\theta)[1 - P_j(\theta)]$ , and the total information from  $n$  identical items for which  $b_j \equiv \theta$  and  $a_j \equiv a$  is simply  $0.7225 na^2$ . Table 1 gives values of  $i$  and  $\sigma_x^2$  in this simple case for selected test lengths and values of  $a$ . Note that where  $1.7a = 1.0$  (i.e.,  $a = .588$ , corresponding to an item trait correlation of .7071 in a standard normal population), four additional items provide a unit gain in precision. The results provide an indication of the amount of information about  $\theta$  that is employed in JML estimation of item parameters. It is apparent that as test length increases, information (i.e., precision) increases at a constant rate and the posterior variance decreases at a decreasing rate.

-----  
 Insert Table 1 about here  
 -----

The magnitude of gain in information about  $\theta$  obtained by assuming an undifferentiated population (i.e.,  $F_1 \equiv F$ ) can be gauged by extending the approximation employed for Table 1. If the normalized likelihood function induced by  $x_1$  is again approximated as  $N(\hat{\theta}, \sigma_x^2)$  and if it is further assumed that examinee 1 has been selected at random from a population in which  $\theta \sim N(\mu, \sigma^2)$ , then

$$p(\theta | x_1) \sim N(\tilde{\theta}, \tilde{\Sigma}) \quad ,$$



where

$$\hat{\theta} = \frac{\hat{\theta}\sigma_x^{-2} + \mu\sigma^{-2}}{\sigma_x^{-2} + \sigma^{-2}}$$

and

$$\tilde{\Sigma} = (\sigma^{-2} + \sigma_x^{-2})^{-1}.$$

Table 2 shows values of the reciprocal of  $\tilde{\Sigma}$  (i.e., "precision") from various test lengths with identical items with  $1.7a = 1$  and a standard normal prior for  $\theta$ . Note that for each test length, a unit gain in precision is achieved over the  $1.7a = 1$  column of Table 1. These tabled values fall within the ranges encountered in applied work, and suggest that the assumed distribution contributes about as much information about  $\theta$  as four additional items. The corresponding value for  $1.7a = .5$  is sixteen items, and that for  $1.7a = 1.5$  is about one item. Since the absolute contribution is constant with respect to increasing test length, the relative contribution declines.

To gauge the additional impact of differentiating the population through auxiliary variables, we may consider numerical

values resulting from a regression model with homoscedastic residuals. Suppose  $y$  values account for  $(100 \times z)$ -percent of the variance in a population with total variance 1.0, so that  $F_y(\theta) \sim N(\mu_y, \sigma_e^2)$  with  $\sigma_e^2 = 1 - r$ . If the normalized likelihood induced by item responses is approximately  $N(\hat{\theta}, \sigma_x^2)$ , then

$$p(\theta | x_1, y_1) \propto N \left[ \frac{\hat{\theta} \sigma_x^{-2} + \mu_y \sigma_e^{-2}}{\sigma_x^{-2} + \sigma_e^{-2}}, (\sigma_e^{-2} + \sigma_x^{-2})^{-1} \right] .$$

Using the same simplified item response model and 'a' value as Table 2, Table 3 compares values of the inverse of the posterior variance for  $\theta$  as determined by (i) item responses alone, (ii) with knowledge of membership in an undifferentiated population with unit variance, and (iii) with the additional knowledge of auxiliary variables that account for successively greater proportions of total variance. Values between 10- and 40-percent, a range typical of educational and psychological work, increase information (posterior precision) about  $\theta$  by amounts roughly equivalent to one to three additional item responses. For items with  $1.7a = .5$ , gains in item units would be doubled; for items with  $1.7a = 1.5$ , gains in item units would be halved.

---

Insert Tables 2-3 about here

---

The Ignorability of  $p(y)$ 

This section demonstrates that under reasonable assumptions, the population distribution of  $y$  can be ignored for the purposes of estimating item parameters  $\beta$  and population parameter  $\alpha$ . Suppose that the distribution of  $y$  in a population of examinees is governed by the density function  $p(y|\gamma)$ , which depends on possibly unknown parameters  $\gamma$  but not upon item parameters  $\beta$  nor on the parameters  $\alpha$  of the conditional distributions  $f(\theta|y, \alpha)$ . The probability of observing the data matrix  $(X, Y)$  from a random sample of  $N$  examinees is given by

$$\begin{aligned}
 P(\underline{X}, \underline{Y} | \underline{\beta}, \underline{\alpha}, \gamma) &= \prod_i \int p(\underline{x}_i | \theta, y_i, \underline{\beta}, \underline{\alpha}, \gamma) p(\theta | y_i, \underline{\beta}, \underline{\alpha}, \gamma) p(y_i | \underline{\beta}, \underline{\alpha}, \gamma) d\theta \\
 &= \prod_i \int p(\underline{x}_i | \theta, \underline{\beta}) p(\theta | y_i, \underline{\alpha}) p(y_i | \gamma) d\theta \\
 &= \left\{ \prod_i \int p(\underline{x}_i | \theta, \underline{\beta}) p(\theta | y_i, \underline{\alpha}) d\theta \right\} \times \left\{ \prod_i p(y_i | \gamma) \right\} \\
 &= P(\underline{X} | \underline{Y}, \underline{\beta}, \underline{\alpha}) P(\underline{Y} | \gamma) \quad .
 \end{aligned} \tag{6}$$

Likelihood inferences about  $\alpha$  and  $\beta$  are therefore independent of inferences about  $\gamma$ , and the conditional MLE's of  $\alpha$  and  $\beta$  given  $\gamma$  are identical to MLE's obtained jointly with  $\gamma$ .

#### Models and Methods

This section presents two IRT models that differentiate examinees by means of auxiliary variables, and suggests computing approximations based on Bock and Aitkin's (1981) marginal maximum likelihood (empirical Bayes) procedures.

#### Mixtures of Finite Distributions

Mislevy (1984) describes a nonparametric approximation of a continuous density function of a latent variable in terms of a distribution with mass at a finite number of prespecified points. The proficiency of each examinee, or  $\theta_1$ , then, is assumed to take one of only  $Q$  known values. The "latent trait" problem is thereby replaced by an analogous "latent class" problem that is easier to solve. A single population was addressed in that presentation, and item parameters were assumed known. We now consider extensions to the simultaneous estimation of item parameters, and to multiple subpopulations indexed by an auxiliary variable  $y$ . This approach provides considerably flexibility in the distributions  $F_1(\theta) = F_{y_1}(\theta)$ . It lends itself well to discrete auxiliary variables with relatively few values.

It proves convenient to write such an auxiliary variable as a vector of 0/1 indicators. Define  $\underline{y}_i = (y_{i1}, \dots, y_{iK})$  by letting  $y_{ik} = 1$  if examinee  $i$  is associated with the  $k$ 'th of  $K$  exhaustive and mutually exclusive subpopulations, and zero otherwise. The probability of observing response pattern  $\underline{x}_i$  from an examinee selected at random from a specified subpopulation is given by

$$p(\underline{x}_i | \underline{y}_i, \beta) = \prod_k \left\{ \int p(\underline{x}_i | \theta, \beta) dF_k(\theta) \right\}^{y_{ik}}, \quad (7)$$

where  $F_k$  is the distribution in subpopulation  $k$ . This probability can be approximated by a finite distribution as

$$p(\underline{x}_i | \underline{y}_i, \beta) \approx \prod_k \left\{ \sum_q p(\underline{x}_i | \theta_q, \beta) W_{qk} \right\}^{y_{ik}} \quad (8)$$

where  $\theta_1, \dots, \theta_Q$  is a grid of points and  $W_{qk}$  is the weight or density at point  $q$  in subpopulation  $k$ . The weights  $W$  play the role of  $\alpha$  in earlier notation. For the remainder of this subsection, we limit our attention to distributions of the form of the right-hand side of Equation 8. As demonstrated above, we may carry out the estimation of  $\beta$  and  $W$  conditional on  $\underline{Y}$ .

Let  $(\underline{X}, \underline{Y})$  be the data matrix observed from a sample of  $N$  examinees selected either randomly from the population as a whole or as random subsamples stratified on  $y$ . The probability of  $\underline{X}$  given  $\underline{Y}$  is proportional to

$$L_M = \prod_i \prod_k \left\{ \sum_q p(\underline{x}_i | \underline{\theta}_q, \underline{\beta}) w_{qk} \right\}^{y_{ik}}$$

and its logarithm is

$$\begin{aligned} l_M &= \log L_M \\ &= \sum_i \sum_k y_{ik} \log \sum_q p(\underline{x}_i | \underline{\theta}_q, \underline{\beta}) w_{qk} \end{aligned}$$

Relative maxima with respect to  $\underline{\beta}$  and  $\underline{W}$  can be obtained by means of the EM algorithm, under the special case of missing indicators for a multinomial distribution (Dempster et al., 1977, Section 4.3). The expectation step of cycle  $t + 1$  computes expected values of the following quantities:

1. The expected number of examinees with proficiency  $\theta_q$  from a sample of size  $N_k$  from subpopulation  $k$ , conditional on  $\underline{X}$ ,  $\underline{Y}$ ,  $\hat{\underline{\beta}}^t$ , and  $\hat{\underline{W}}^t$ :

$$\hat{N}_{qk}^{t+1} = \sum_i y_{ik} \hat{p}_k^t(\theta_q | \underline{x}_i) \quad ,$$

where

$$\hat{p}_k^t(\theta_q | \underline{x}_i) = p(\underline{x}_i | \theta_q, \underline{\beta} = \hat{\underline{\beta}}^t) \hat{W}_{qk}^t / \sum_r p(\underline{x}_i | \theta_r, \underline{\beta} = \hat{\underline{\beta}}^t) \hat{W}_{rk}^t \quad ,$$

an application of Bayes theorem, gives the posterior probability that the proficiency of examinee  $i$  is  $\theta_q$ , given provisional parameter estimates  $\hat{\underline{\beta}}^t$  and  $\hat{\underline{W}}^t$ .

2. The expected number of correct responses to item  $j$  from examinees in subpopulation  $k$  with proficiency  $\theta_q$ , given a random sample of size  $N_k$  (again given  $\hat{\underline{\beta}}^t$  and  $\hat{\underline{W}}^t$ ):

$$\hat{R}_{jqk}^{t+1} = \sum_i y_{ik} x_{ij} \hat{p}_k^t(\theta_q | \underline{x}_i) \quad .$$

The maximization step computes what would be MLE's of  $\beta$  and  $\tilde{W}$  if  $\hat{\tilde{N}}$  and  $\hat{\tilde{R}}$  were observed quantities rather than conditional expectations. For  $\tilde{W}$ , we have simply

$$\hat{\tilde{W}}_{qk}^{t+1} = \hat{\tilde{N}}_{qk}^{t+1} / N_k \quad .$$

For  $\beta$ , we solve conditional expectations of likelihood equations:

$$0 = \sum_q \frac{\hat{\tilde{R}}_{jq+}^{t+1} \hat{\tilde{N}}_{q+}^{t+1} P_j(\theta_q)}{P_j(\theta_q) [1 - P_j(\theta_q)]} \frac{\partial P_j(\theta_q)}{\partial \beta} \quad , \quad (9)$$

where  $\hat{\tilde{R}}_{jq+}^{t+1} = \sum_k \hat{\tilde{R}}_{jqk}^{t+1}$  and  $\hat{\tilde{N}}_{q+}^{t+1}$  is similarly defined. Under the 2-parameter logistic model, for example, Equation 9 simplifies as follows:

$$a_j: \quad 0 = \sum_q [\hat{\tilde{R}}_{jq+}^{t+1} - \hat{\tilde{N}}_{q+}^{t+1} P_j(\theta_q)] (\theta_q - b_j)$$

$$b_j: \quad 0 = \sum_q [\hat{\tilde{R}}_{jq+}^{t+1} - \hat{\tilde{N}}_{q+}^{t+1} P_j(\theta_q)] a_j \quad .$$

In principle, the linear indeterminacy in the 1-, 2-, and 3-parameter logistic and normal IRT models presents no impediment to



the EM algorithm, which readily converges to one of the infinitely many solutions on a ridge. Numerical stability and the quality of the finite characterization of  $F$  are enhanced, however, by controlling the scaling of the solution at this point. One convenient way of doing so is to standardize the weighted average distribution. We have referred to the points  $\theta_q$  as specified a priori; given the linear indeterminacy, we may conceive of only their relative spacing as prespecified. After each EM cycle, then, we may rescale the points as follows:

$$\theta_q = (\theta_q - \bar{\theta})/s$$

where

$$\bar{\theta} = N^{-1} \sum_k N_k \sum_q \theta_q \hat{w}_{qk}^t$$

and

$$s = N^{-1} \sum_k N_k \sum_q (\theta_q - \bar{\theta})^2 \hat{w}_{qk}^t .$$

Item parameters are adjusted accordingly. Under the 2- and 3-

parameter models,  $\hat{b}_j$  is replaced by  $(\hat{b}_j - \bar{O})/s$  and  $a_j$  is replaced by  $sa_j$ . Under 1-parameter models, rescaling takes place only with respect to  $\bar{O}$ .

Iteration from several starting values helps to verify whether a given solution is indeed a global maximum. The observed information matrix for the item parameter estimates can then be approximated via Equation 6. Employing Louis's (1982) simplifications for "missing multinomial indicators" problems, we obtain

$$I_{X,Y}(\hat{\beta}) \approx \sum_i \sum_k y_{ik} \sum_q \left( \frac{\partial \ell_1(\theta_q)}{\partial \beta} \right) \left( \frac{\partial \ell_1(\theta_q)}{\partial \beta} \right)' \hat{p}_k(\theta_q | \underline{x}_i), \quad (10)$$

where  $\hat{p}_k(\theta_q | \underline{x}_i)$  is evaluated at  $\hat{\beta}$  and  $\hat{W}$ .

#### A Linear Model

The unrestricted mixture solution described above becomes unwieldy as the number of potential values of the auxiliary variable increases. The more structured alternative of a linear model for  $p(\theta | y)$  is suitable when  $y$  is vector-valued or is continuo rather than discrete. Assuming homoscedastic and normal residuals, we would have

$$\theta \sim N(\underline{y}'\underline{\alpha}, \sigma^2) \quad ,$$

where auxiliary variables are coded so that the  $K$  columns of  $\underline{Y} = (\underline{y}_1 : \dots : \underline{y}_K)'$ , which are basis vectors for the  $K$  elements of  $\underline{\alpha}$ , are linearly independent. They may include values on measured variables such as previous test scores and dummy regression variables that encode selected contrasts among categorical auxiliary variables.

Maximum likelihood solutions for  $\underline{\alpha}$  and  $\sigma^2$  in the special case of structured means for the cells of a multi-way design have been given by Mislevy (1985) under the assumption that item parameters are known, and by Zwarts and Veldhuesen (1985) under the assumption that  $p(\underline{x}|\theta)$  is the Rasch model with unknown item parameters to be estimated jointly. These solutions are readily extended to the case of a general IRT model with unknown item parameters. This section describes an approximation over a grid of prespecified points so that computation is similar to the nonparametric solution described above. Attention is focused for convenience upon the 1-, 2-, and 3-parameter logistic and normal IRT models.

The linear indeterminacies of these models are again conveniently resolved by restrictions on the population parameters. First, we may without loss of generality fix  $\sigma^2$  at unity to set the

unit-size of the scale. For 1-parameter models, a slope parameter common over items is then estimated. Second, we may set the origin by centering the elements of each column of  $\tilde{Y}$  at zero. All effects are thus cast as deviations around a grand mean of zero. This restriction, in conjunction with the independence of the basis vectors, completes the resolution of the scale.

The marginal likelihood for a sample of size  $N$  is written as

$$L = \prod_i \int p(\tilde{x}_i | \theta, \beta) \phi(\theta - \tilde{y}_i' \alpha) d\theta, \quad ,$$

where  $\phi$  represents the standard normal density function.

Approximation over a finite grid of points is accomplished by

$$L^* = \prod_i \sum_q p(\tilde{x}_i | \theta_q, \beta) w_{qi}(\alpha), \quad ,$$

where

$$w_{qi}(\alpha) = \exp[-(\theta_q - \tilde{y}_i' \alpha)^2 / 2] / \sum_r \exp[-(\theta_r - \tilde{y}_i' \alpha)^2 / 2] \quad .$$

The weights  $w$  play the same role as those in the preceding approximation. The difference is that they are no longer estimated

without restriction, but modeled as functions of the effect parameters  $\alpha$ .

MML estimation can again proceed in EM cycles that solve the likelihood equations. Let  $\hat{\beta}^t$  and  $\hat{\alpha}^t$  be provisional estimates from cycle  $t$ . The E-step computes expected counts of examinees and correct responses at each point:

$$\hat{N}_q^{t+1} = \sum_i P(\theta_q | x_i, \hat{\beta}^t, \hat{\alpha}^t)$$

and

$$\hat{R}_{jq}^{t+1} = \sum_i x_{ij} P(\theta_q | x_i, \hat{\beta}^t, \hat{\alpha}^t)$$

where

$$P(\theta_q | x_i, \hat{\beta}^t, \hat{\alpha}^t) = p(x_i | \theta_q, \hat{\beta}^t) w_{qi}(\hat{\alpha}^t) / \sum_r p(x_i | \theta_r, \hat{\beta}^t) w_{ri}(\hat{\alpha}^t) .$$

It also computes the conditional expected value of each examinee's proficiency:

$$\hat{\theta}_i^{t+1} = \sum_q \theta_q P(\theta_q | x_i, \hat{\beta}^t, \hat{\alpha}^t) .$$

The M-step pseudo-likelihood equations for item parameters can be written as in Equation 9. The equations for  $\alpha$  simplify to

$$\hat{\alpha}^{t+1} = (Y'Y)^{-1} Y' \hat{\theta}^{t+1},$$

where  $\hat{\theta}^{t+1} = (\hat{\theta}_1^{t+1}, \dots, \hat{\theta}_N^{t+1})$ . The posterior information matrix for  $\hat{\beta}$  can again be approximated via Equation 10.

#### A Numerical Example

This section illustrates the procedures described above. The data are responses to four items from the Arithmetic Reasoning test of the Armed Services Vocational Aptitude Battery (ASVAB), Form 8A, as observed in a sample of 776 participants in the Profile of American Youth survey (U.S. Department of Defense, 1982). Table 4 gives counts of the sixteen possible response patterns occurring in each cell of a 2-by-2 design based on two background variables collected along with item responses. Because these variables are based on demographic information rather than the educationally-relevant information we would prefer, we will refer to the factors as simply Factor A and Factor B, nesting levels 1 and 2 within each.

-----  
Insert Table 4 about here  
-----

Four analyses were carried out on these data. In each, the 2-parameter logistic ogive was employed as the IRT model for conditional probabilities of correct response. The analyses differed in terms of the auxiliary information about examinees they employed. The first run used MML estimation of item parameters and densities over a grid of ten points, assuming examinees were drawn at random from a single undifferentiated population. The second and third runs differentiated the population via Factor A and Factor B respectively, and the fourth run employed both factors jointly.

Resulting item parameter estimates and standard errors, along with subpopulation means and standard deviations, are shown in Tables 5 through 8. The scale has been set in all solutions to standardize the total population. For each item parameter type, columns in Table 6 through 8 display the ratio of the squared standard error of the item parameter estimate under the undifferentiated model to the corresponding value in the differentiated model. The result can be interpreted as efficiency relative to the undifferentiated model, and the excess of a value above unity reflects the proportional increase in estimation precision. Geometric averages are also shown for the relative efficiency columns. The excess of such a value over unity, times

four, gives the increases of precision in the units of numbers of additional items of the same kind.

-----  
Insert Tables 5-8 about here  
-----

It is apparent that including auxiliary information had little effect on the values of the item parameter estimates. The differences between the estimates from the undifferentiated and the fully differentiated solutions occur only in the second decimal place. More significant differences exist in the accompanying (estimated) standard errors, however. The precision of threshold estimates was improved only modestly; an increase roughly equivalent to one additional item response per examinee was observed in the fully differentiated run. The precision of slope estimates was improved dramatically; an increase roughly equivalent to eight items was observed. It would appear that Factor A accounted for more increase in precision for slopes, while Factor B accounted for more increase in precision for thresholds.

#### Discussion

This paper has outlined procedures for incorporating auxiliary information about examinees into the IRT framework. Enhancing the precision of item parameter estimates was the primary focus. This section evaluates the value of improvements so attained, and discusses two additional aspects of the model.



The increase in information about item parameters in typical educational and psychological settings can be expected to lie in the range of two to six items. The numerical example suggests that the increase will vary by item parameter type, probably less for well-estimated parameters and greater for poorly-estimated parameters.

The expected increase is modest, to be sure, but in many applications it is free in the sense that it is already available for use. Because its incremental value decreases for longer tests, auxiliary information would be most useful in settings where relatively few response are solicited from each examinee. This would include two applications of great current interest, namely educational assessment and adaptive testing. In assessment, data that are sparse at the level of individuals--say, five items in a given scale--yield more efficient estimates of population parameters for a given total number of item responses. In adaptive testing, new items are calibrated using joint response patterns with previously-calibrated items while the number of old items is held to minimally acceptable levels--as few as, say, fifteen.

A side issue in the present paper but a fundamentally important result is that when examinees are indeed a random sample from a well-defined population, the estimated population

distributions and effect parameters are consistent within the limits of precision afforded by the numerical approximations (see Mislevy, 1984, 1985, on population estimation when item parameters are known). This stands in contrast to the asymptotically biased results obtained by using the distribution of  $\hat{\theta}$  to approximate the distribution of  $\theta$ . In fact, the discrepancy between the two distributions is largest in exactly those cases in which the present procedures offer most the benefit for item parameter estimation, namely short tests.

Finally, it is implicit in preceding discussions that auxiliary information about examinees can lead to improved estimates of individual proficiencies. Whether estimates that are improved in the sense of minimum mean squared error are unequivocally "better" for all applications is not clear, however. We have avoided advocating the use of auxiliary information when tests are used as contests--i.e., when important placement or selection decisions are made for individual examinees--because it would seem that in these situations the tester ought to gather enough data directly dependent upon proficiency (i.e., item responses) to make satisfactorily precise decisions on that strength alone. In adaptive testing, for example, we would recommend the use of auxiliary information to improve item parameter estimation, but not to estimate scores that will be used to compare individual examinees.

## References

- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F M. Lord and M. R. Novick, Statistical theories of mental test scores (pp. 397-479). Reading, MA: Addison-Wesley.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. Psychometrika, 46, 443-459.
- Bock, R. D., & Lieberman, M. (1970). Fitting a response model for n dichotomously scores items. Psychometrika, 35, 179-197.
- de Leeuw, J., & Verhelst, N. (1984). Maximum likelihood estimation in generalized Rasch models (Department of Data Theory Research Report RR-84-11). Leiden, The Netherlands: University of Leiden.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from complete data via the EM algorithm (with discussion. Journal of the Royal Statistical Society, Series B, 39, 1-38.
- Leonard, T., & Novick, M. R. (1985). Bayesian inference and diagnostics for the three-parameter logistic model (ONR Technical Report 85-5). Iowa City, IA: CADA Research Group, University of Iowa.

- Lindley, D. V., & Smith A. F. M. (1972). Bayes estimates for the linear model. Journal of the Royal Statistical Society, Series B, 34, 1-41.
- Louis, T. (1982). Finding the observed information matrix when using the EM algorithm. Journal of the Royal Statistical Society, Series B, 44, 226-233.
- Mislevy, R. J. (1984). Estimating latent distributions. Psychometrika, 49, 359-381.
- Mislevy, R. J. (1985). Estimate of latent group effects. Journal of the American Statistical Association, 80, 993-997.
- Rigdon, S. E., & Tsutakawa, R. K. (1982). Parameter estimation in latent trait models. Psychometrika, 48, 567-574.
- Swaminathan, H., & Gifford, J. A. (1982). Bayesian estimation in the Rasch model. Journal of Educational Statistics, 9, 23-80.
- Tjur, T. (1982). A connection between Rasch's item analysis model and a multiplicative Poisson model. Scandinavian Journal of Statistics, 9, 23-30.
- U.S. Department of Defense (1982). Profile of American youth Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics).
- Zacks, S. (1971). The theory of statistical inference. New York: Wiley.

Zwarts, M., & Veldhuesen, N. (1985). Maximum marginal likelihood estimation in latent trait models. Paper presented at the European meeting of the Psychometric Society, Cambridge, England.

Acknowledgment

This work was supported by contract no. N00014-85-K-0683, project designation NR 150-539, from Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research. The author is grateful to Kathleen Sheehan and Martha Stocking for their comments and suggestions.

Table 1  
Posterior Precision for  $\theta$  from Item Responses Only

n	<u>1.7a = .500</u>		<u>1.7a = 1.000</u>		<u>1.7a = 1.500</u>	
	i	$\sigma_x^2$	i	$\sigma_x^2$	i	$\sigma_x^2$
2	.125	8.000	.500	2.000	1.125	.889
4	.250	4.000	1.000	1.000	2.250	.444
8	.500	2.000	2.000	.500	4.500	.222
16	1.000	1.000	4.000	.250	9.000	.111
32	2.000	.500	8.000	.125	18.000	.056
64	4.000	.250	16.000	.063	36.000	.028
128	8.000	.125	32.000	.031	72.000	.014

n = number of identical items with a as noted and b =  $\theta$ .

i = information  $\equiv$  posterior precision.

Table 2

Posterior Precision for  $\theta$  from Item Responses and Population Membership

$$1.7a = 1.000$$

n	$i(\equiv (\Sigma^{-1}))$	Relative Efficiency ( $\sigma_x^2/\Sigma$ )	Effective Gain
2	1.500	3.000	200.0%
4	2.000	2.000	100.0%
8	3.000	1.500	50.0%
16	5.000	1.250	25.0%
32	9.000	1.125	12.5%
64	17.000	1.063	6.3%
128	33.000	1.031	3.1%

$n$  = number of identical items with  $a$  as noted and  $b = \theta$ .

$i$  = information  $\equiv$  posterior precision.



Table 3

Precision Increases for  $\theta$  Resulting from the Use of  
Auxiliary Information

Source	Increment in Posterior Precision	Precision Gain in Item Units	Gain over Undifferentiated Population
One-item response	.250	1.000	--
Population membership	1.000	4.000	--
Auxiliary information			
$R^2 = .10$	1.111	4.444	11.1%
$R^2 = .20$	1.250	5.000	25.0%
$R^2 = .30$	1.429	5.716	42.9%
$R^2 = .40$	1.667	6.668	66.7%
$R^2 = .50$	2.000	8.000	100.0%
$R^2 = .60$	2.500	10.000	150.0%
$R^2 = .70$	3.333	13.332	233.3%
$R^2 = .80$	5.000	20.000	400.0%
$R^2 = .90$	10.000	40.000	900.0%

Table 4  
Counts of Observed Response Patterns

Item Response				A1B1	A1B2	A2B1	A2B2
1	2	3	4				
0	0	0	0	23	20	27	29
0	0	0	1	5	8	5	8
0	0	1	0	12	14	15	7
0	0	1	1	2	2	3	3
0	1	0	0	16	20	16	14
0	1	0	1	3	5	5	5
0	1	1	0	6	11	4	6
0	1	1	1	1	7	3	0
1	0	0	0	22	23	15	14
1	0	0	1	6	8	10	10
1	0	1	0	7	9	8	11
1	0	1	1	19	6	1	2
1	1	0	0	21	18	7	19
1	1	0	1	11	15	9	5
1	1	1	0	23	20	10	8
1	1	1	1	86	42	2	4
Total				263	228	140	145

Table 5

Item Parameter Estimates: Undifferentiated Population

Item	$\hat{b}$	$SE(\hat{b})$	$\hat{a}$	$SE(a)$
1	-.422	.058	1.022	.171
2	-.226	.072	.666	.094
3	.152	.076	.705	.096
4	.397	.080	.839	.114
Population Mean:		0.000		
Population Standard Deviation:		1.000		

Table 6

Item Parameter Estimates: Population Differentiated  
with Respect to Factor A Only

Item	$\hat{b}$	SE( $\hat{b}$ )	Relative Efficiency	$\hat{a}$	SE( $\hat{a}$ )	Relative Efficiency
1	-.436	.062	.875	.869	.069	6.142
2	-.217	.077	.874	.622	.054	3.030
3	.189	.072	1.114	.676	.056	2.939
4	.465	.069	1.344	.775	.061	3.493
<hr/>						
Geometric average						
relative efficiency:			1.035			3.718
Subpopulation means:				.296, -.511		
Subpopulation standard deviations:				.960, .850		

Table 7

Item Parameter Estimates: Population Differentiated  
with Respect to Factor B Only

Item	$\hat{b}$	SE( $\hat{b}$ )	Relative Efficiency	$\hat{a}$	SE( $\hat{a}$ )	Relative Efficiency
1	-.408	.057	1.035	.941	.073	5.487
2	-.211	.077	.874	.621	.056	2.818
3	.185	.071	1.146	.686	.058	2.740
4	.431	.064	1.563	.842	.067	2.895
<hr/>						
Geometric average relative efficiency:			1.128	3.328		
Subpopulation means:			.136, -.147			
Subpopulation standard deviations:			1.021, .955			

Table 8

Item Parameter Estimates: Population Differentiated  
with Respect to Factors A and B

Item	$\hat{b}$	SE( $\hat{b}$ )	Relative Efficiency	$\hat{a}$	SE( $\hat{a}$ )	Relative Efficiency
1	-.421	.052	1.244	1.006	.080	4.569
2	-.213	.071	1.028	.672	.059	2.538
3	.139	.065	1.367	.775	.063	2.311
4	.402	.066	1.469	.834	.066	2.983
Geometric average relative efficiency:			1.266	2.994		
Subpopulation means:				.485, .073, -.513, -.502		
Subpopulation standard deviations:				1.164, .855, .642, .640		

## Educational Testing Service/Mislevy

Personnel Analysis Division,  
AF/MPXA

5C360, The Pentagon  
Washington, DC 20330

Air Force Human Resources Lab  
AFHRL/MPD

Brooks AFB, TX 78235

Dr. Earl A. Alluisi  
HQ, AFHRL (AFSC)  
Brooks AFB, TX 78235

Dr. Erling B. Andersen  
Department of Statistics  
Studiestraede 6  
1455 Copenhagen  
DENMARK

Dr. Phipps Arabie  
University of Illinois  
Department of Psychology  
603 E. Daniel St.  
Champaign, IL 61820

Technical Director, ARI  
5001 Eisenhower Avenue  
Alexandria, VA 22333

Dr. Eva L. Baker  
UCLA Center for the Study  
of Evaluation  
145 Moore Hall  
University of California  
Los Angeles, CA 90024

Dr. Isaac Bejar  
Educational Testing Service  
Princeton, NJ 08450

Dr. Menucha Birenbaum  
School of Education  
Tel Aviv University  
Tel Aviv, Ramat Aviv 69978  
ISRAEL

Dr. Arthur S. Blaiwes  
Code N711  
Naval Training Equipment Center  
Orlando, FL 32813

Dr. Bruce Bloxom  
Administrative Sciences  
Code 54B1  
Navy Postgraduate School  
Monterey, CA 93943

Dr. R. Darrell Bock  
University of Chicago  
Department of Education  
Chicago, IL 60637

Cdt. Arnold Bohrer  
Sectie Psychologisch Onderzoek  
Rekruterings-En Selectiecentrum  
Kwartier Koningen Astrid  
Bruijnstraat  
1120 Brussels, BELGIUM

Dr. Robert Breaux  
Code N-095R  
NAVTRAEQUIPCEN  
Orlando, FL 32813

Dr. Robert Brennan  
American College Testing  
Programs  
P. O. Box 168  
Iowa City, IA 52243

Dr. Patricia A. Butler  
NIE Mail Stop 1806  
1200 19th St., NW  
Washington, DC 20208

Mr. James W. Carey  
Commandant (G-PTE)  
U.S. Coast Guard  
2100 Second Street, S.W.  
Washington, DC 20593

Dr. James Carlson  
American College Testing  
Program  
P.O. Box 168  
Iowa City, IA 52243

Dr. John B. Carroll  
409 Elliott Rd.  
Chapel Hill, NC 27514

Dr. Robert Carroll  
NAVOP 01B7  
Washington, DC 20370

## Educational Testing Service/Mislevy

Dr. Norman Cliff  
Department of Psychology  
Univ. of So. California  
University Park  
Los Angeles, CA 90007

Director,  
Manpower Support and  
Readiness Program  
Center for Naval Analysis  
2000 North Beauregard Street  
Alexandria, VA 22311

Dr. Stanley Collyer  
Office of Naval Technology  
Code 272  
800 N. Quincy Street  
Arlington, VA 22217-5000

Dr. Hans Crombag  
University of Leyden  
Education Research Center  
Boerhaavelaan 2  
2334 EN Leyden  
The NETHERLANDS

CTB/McGraw-Hill Library  
2500 Garden Road  
Monterey, CA 93940

Dr. Dattprasad Divgi  
Center for Naval Analysis  
4401 Ford Avenue  
P.O. Box 16268  
Alexandria, VA 22302-0268

Dr. Hei-Ki Dong  
Ball Foundation  
800 Roosevelt Road  
Building C, Suite 206  
Glen Ellyn, IL 60137

Defense Technical  
Information Center  
Cameron Station, Bldg 5  
Alexandria, VA 22314  
Attn: TC  
(12 Copies)

Dr. Stephen Dunbar  
Lindquist Center  
for Measurement  
University of Iowa  
Iowa City, IA 52242

Dr. James A. Earles  
Air Force Human Resources Lab  
Brooks AFB, TX 78235

Dr. Kent Eaton  
Army Research Institute  
5001 Eisenhower Avenue  
Alexandria, VA 22333

Dr. John M. Eddins  
University of Illinois  
252 Engineering Research  
Laboratory  
103 South Mathews Street  
Urbana, IL 61801

Dr. Susan Embretson  
University of Kansas  
Psychology Department  
Lawrence, KS 66045

ERIC Facility-Acquisitions  
4833 Rugby Avenue  
Bethesda, MD 20014

Dr. Benjamin A. Fairbank  
Performance Metrics, Inc.  
5825 Callaghan  
Suite 225  
San Antonio, TX 78228

Dr. Leonard Feldt  
Lindquist Center  
for Measurement  
University of Iowa  
Iowa City, IA 52242

Dr. Richard L. Ferguson  
American College Testing  
Program  
P.O. Box 168  
Iowa City, IA 52240

Dr. Gerhard Fischer  
Liebiggasse 5/3  
A 1010 Vienna  
AUSTRIA



## Educational Testing Service/Mislevy

Prof. Donald Fitzgerald  
University of New England  
Department of Psychology  
Armidale, New South Wales 2351  
AUSTRALIA

Mr. Paul Foley  
Navy Personnel R&D Center  
San Diego, CA 92152

Dr. Carl H. Frederiksen  
McGill University  
3700 McTavish Street  
Montreal, Quebec H3A 1Y2  
CANADA

Dr. Robert D. Gibbons  
University of Illinois-Chicago  
P.O. Box 6998  
Chicago, IL 69680

Dr. Janice Gifford  
University of Massachusetts  
School of Education  
Amherst, MA 01003

Dr. Robert Glaser  
Learning Research  
& Development Center  
University of Pittsburgh  
3929 O'Hara Street  
Pittsburgh, PA 15260

Dr. Bert Green  
Johns Hopkins University  
Department of Psychology  
Charles & 34th Street  
Baltimore, MD 21218

Dr. Ronald K. Hambleton  
Prof. of Education & Psychology  
University of Massachusetts  
at Amherst  
Hills House  
Amherst, MA 01003

Ms. Rebecca Hetter  
Navy Personnel R&D Center  
Code 62  
San Diego, CA 92152

Dr. Paul W. Holland  
Educational Testing Service  
Rosedale Road  
Princeton, NJ 08541

Prof. Lutz F. Hornke  
Universitat Dusseldorf  
Erziehungswissenschaftliches  
Universitätsstr. 1  
Dusseldorf 1  
WEST GERMANY

Dr. Paul Horst  
677 G Street, #184  
Chula Vista, CA 90010

Mr. Dick Hoshaw  
NAVOP-135  
Arlington Annex  
Room 2834  
Washington, DC 20350

Dr. Lloyd Humphreys  
University of Illinois  
Department of Psychology  
603 East Daniel Street  
Champaign, IL 61820

Dr. Steven Hunka  
Department of Education  
University of Alberta  
Edmonton, Alberta  
CANADA

Dr. Huynh Huynh  
College of Education  
Univ. of South Carolina  
Columbia, SC 29208

Dr. Robert Jannarone  
Department of Psychology  
University of South Carolina  
Columbia, SC 29208

Dr. Douglas H. Jones  
Advanced Statistical  
Technologies Corporation  
10 Trafalgar Court  
Lawrenceville, NJ 08148

1986/05/04

Educational Testing Service/Mislevy

Dr. G. Gage Kingsbury  
Portland Public Schools  
Research and Evaluation Department  
501 North Dixon Street  
P. O. Box 3107  
Portland, OR 97209-3107

Dr. William Koch  
University of Texas-Austin  
Measurement and Evaluation  
Center  
Austin, TX 78703

Dr. Leonard Kroeker  
Navy Personnel R&D Center  
San Diego, CA 92152

Dr. Michael Levine  
Educational Psychology  
210 Education Bldg.  
University of Illinois  
Champaign, IL 61801

Dr. Charles Lewis  
Faculteit Sociale Wetenschappen  
Rijksuniversiteit Groningen  
Oude Boteringestraat 23  
9712GC Groningen  
The NETHERLANDS

Dr. Robert Linn  
College of Education  
University of Illinois  
Urbana, IL 61801

Dr. Robert Lockman  
Center for Naval Analysis  
4401 Ford Avenue  
P.O. Box 16268  
Alexandria, VA 22302-0268

Dr. Frederic M. Lord  
Educational Testing Service  
Princeton, NJ 08541

Dr. James Lumsden  
Department of Psychology  
University of Western Australia  
Nedlands W.A. 6009  
AUSTRALIA

Dr. William L. Maloy  
Chief of Naval Education  
and Training  
Naval Air Station  
Pensacola, FL 32508

Dr. Gary Marco  
Stop 31-E  
Educational Testing Service  
Princeton, NJ 08451

Dr. Clessen Martin  
Army Research Institute  
5001 Eisenhower Blvd.  
Alexandria, VA 22333

Dr. James McBride  
Psychological Corporation  
c/o Harcourt, Brace,  
Javanovich Inc.  
1250 West 6th Street  
San Diego, CA 92101

Dr. Clarence McCormick  
HQ, MEPCOM  
MEPCT-P  
2500 Green Bay Road  
North Chicago, IL 60064

Mr. Robert McKinley  
University of Toledo  
Department of Educational Psychology  
Toledo, OH 43606

Dr. Barbara Means  
Human Resources  
Research Organization  
1100 South Washington  
Alexandria, VA 22314

Dr. Robert Mislevy  
Educational Testing Service  
Princeton, NJ 08541

Headquarters, Marine Corps  
Code MPI-20  
Washington, DC 20380

Dr. W. Alan Nicewander  
University of Oklahoma  
Department of Psychology  
Oklahoma City, OK 73069

## Educational Testing Service/Mislevy

Dr. William E. Nordbrock  
FMC-ADCO Box 25  
APO, NY 09710

Dr. Melvin R. Novick  
356 Lindquist Center  
for Measurement  
University of Iowa  
Iowa City, IA 52242

Director, Manpower and Personnel  
Laboratory,  
NPRDC (Code 06)  
San Diego, CA 92152

Library, NPRDC  
Code P201L  
San Diego, CA 92152

Commanding Officer,  
Naval Research Laboratory  
Code 2627  
Washington, DC 20390

Dr. James Olson  
WICAT, Inc.  
1875 South State Street  
Orem, UT 84057

Office of Naval Research,  
Code 1142PT  
800 N. Quincy Street  
Arlington, VA 22217-5000  
(6 Copies)

Special Assistant for Marine  
Corps Matters,  
ONR Code OOMC  
800 N. Quincy St.  
Arlington, VA 22217-5000

Dr. Judith Orasanu  
Army Research Institute  
5001 Eisenhower Avenue  
Alexandria, VA 22333

Wayne M. Patience  
American Council on Education  
GED Testing Service, Suite 20  
One Dupont Circle, NW  
Washington, DC 20036

Dr. James Paulson  
Department of Psychology  
Portland State University  
P.O. Box 751  
Portland, OR 97207

Dr. Roger Pennell  
Air Force Human Resources  
Laboratory  
Lowry AFB, CO 80230

Dr. Mark D. Reckase  
ACT  
P. O. Box 168  
Iowa City, IA 52243

Dr. Malcolm Ree  
AFHRL/MP  
Brooks AFB, TX 78235

Dr. Carl Ross  
CNET-PDCD  
Building 90  
Great Lakes NTC, IL 60088

Dr. J. Ryan  
Department of Education  
University of South Carolina  
Columbia, SC 29208

Dr. Fumiko Samejima  
Department of Psychology  
University of Tennessee  
Knoxville, TN 37916

Mr. Drew Sands  
NPRDC Code 62  
San Diego, CA 92152

Dr. Robert Sasmor  
HQDA DAMA-ARL  
Pentagon, Room 3E516  
Washington, DC 20310-0631  
USA

Dr. Mary Schratz  
Navy Personnel R&D Center  
San Diego, CA 92152

Dr. W. Steve Sellman  
OASD(MRA&L)  
2B269 The Pentagon  
Washington, DC 20301

## Educational Testing Service/Mislevy

Dr. Kazuo Shigemasu  
7-9-24 Kugenuma-Kaigan  
Fujusawa 251  
JAPAN

Dr. William Sims  
Center for Naval Analysis  
4401 Ford Avenue  
P.O. Box 16268  
Alexandria, VA 22302-0268

Dr. H. Wallace Sinaiko  
Manpower Research  
and Advisory Services  
Smithsonian Institution  
801 North Pitt Street  
Alexandria, VA 22314

Dr. Richard Sorensen  
Navy Personnel R&D Center  
San Diego, CA 92152

Dr. Paul Speckman  
University of Missouri  
Department of Statistics  
Columbia, MO 65201

Dr. Martha Stocking  
Educational Testing Service  
Princeton, NJ 08541

Dr. Peter Stolloff  
Center for Naval Analysis  
200 North Beauregard Street  
Alexandria, VA 22311

Dr. William Stout  
University of Illinois  
Department of Mathematics  
Urbana, IL 61801

Maj. Bill Strickland  
AF/MPXOA  
4E168 Pentagon  
Washington, DC 20330

Dr. Hariharan Swaminathan  
Laboratory of Psychometric and  
Evaluation Research  
School of Education  
University of Massachusetts  
Amherst, MA 01003

Mr. Brad Sympson  
Navy Personnel R&D Center  
San Diego, CA 92152

Dr. Kikumi Tatsuoka  
CERL  
252 Engineering Research  
Laboratory  
Urbana, IL 61801

Dr. Maurice Tatsuoka  
220 Education Bldg  
1310 S. Sixth St.  
Champaign, IL 61820

Dr. David Thissen  
Department of Psychology  
University of Kansas  
Lawrence, KS 66044

Mr. Gary Thomasson  
University of Illinois  
Educational Psychology  
Champaign, IL 61820

Dr. Robert Tsutakawa  
The Fred Hutchinson  
Cancer Research Center  
Division of Public Health Sci.  
1124 Columbia Street  
Seattle, WA 98104

Dr. Ledyard Tucker  
University of Illinois  
Department of Psychology  
603 E. Daniel Street  
Champaign, IL 61820

Dr. Vern W. Urry  
Personnel R&D Center  
Office of Personnel Management  
1900 E. Street, NW  
Washington, DC 20415

Dr. David Vale  
Assessment Systems Corp.  
2233 University Avenue  
Suite 310  
St. Paul, MN 55114

Dr. Frank Vicino  
Navy Personnel R&D Center  
San Diego, CA 92152

1986/05/04

Educational Testing Service/Mislevy

Dr. Howard Wainer  
Division of Psychological Studies  
Educational Testing Service  
Princeton, NJ 08541

Dr. Ming-Mei Wang  
Lindquist Center  
for Measurement  
University of Iowa  
Iowa City, IA 52242

• Dr. Thomas A. Warm  
Coast Guard Institute  
P. O. Substation 18  
Oklahoma City, OK 73169

Dr. Brian Waters  
Program Manager  
Manpower Analysis Program  
HumRRO  
1100 S. Washington St.  
Alexandria, VA 22314

Dr. David J. Weiss  
N660 Elliott Hall  
University of Minnesota  
75 E. River Road  
Minneapolis, MN 55455

Dr. Ronald A. Weitzman  
NPS, Code 54Wz  
Monterey, CA 92152

Major John Welsh  
AFHRL/MOAN  
Brooks AFB, TX 78223

Dr. Rand R. Wilcox  
University of Southern  
California  
Department of Psychology  
Los Angeles, CA 90007

German Military Representative  
ATTN: Wolfgang Wildegrube  
Streitkraefteamt  
D-5300 Bonn 2  
4000 Brandywine Street, NW  
Washington, DC 20016

Dr. Bruce Williams  
Department of Educational  
Psychology  
University of Illinois  
Urbana, IL 61801

Dr. Hilda Wing  
Army Research Institute  
5001 Eisenhower Ave.  
Alexandria, VA 22333

Dr. Martin F. Wiskoff  
Navy Personnel R & D Center  
San Diego, CA 92152

Mr. John H. Wolfe  
Navy Personnel R&D Center  
San Diego, CA 92152

Dr. George Wong  
Biostatistics Laboratory  
Memorial Sloan-Kettering  
Cancer Center  
1275 York Avenue  
New York, NY 10021

Dr. Wendy Yen  
CTB/McGraw Hill  
Del Monte Research Park  
Monterey, CA 93940

END

DTIC

8-86